

# VU Research Portal

## Editing and Estimation of Measurement Errors in Administrative and Survey Data

Scholtus, S.

2018

### **document version**

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

### **citation for published version (APA)**

Scholtus, S. (2018). *Editing and Estimation of Measurement Errors in Administrative and Survey Data*. [PhD-Thesis - Research and graduation internal, Vrije Universiteit Amsterdam].

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

### **E-mail address:**

[vuresearchportal.ub@vu.nl](mailto:vuresearchportal.ub@vu.nl)

# Table of contents

<b>1</b>	<b>Setting the Problem</b>	<b>9</b>
1.1	Introduction . . . . .	9
1.2	Types of data and types of errors . . . . .	11
1.2.1	Survey data and administrative data . . . . .	11
1.2.2	Errors in statistics . . . . .	15
1.2.3	True scores and true values . . . . .	18
1.2.4	Measurement levels . . . . .	19
1.3	The editing approach . . . . .	20
1.4	The estimation approach . . . . .	23
1.5	Outline of the rest of this thesis . . . . .	28
<b>2</b>	<b>Editing and Estimation of Measurement Errors</b>	<b>33</b>
2.1	Introduction . . . . .	33
2.2	The editing approach . . . . .	33
2.2.1	Methods for statistical data editing . . . . .	33
2.2.2	Administrative data . . . . .	38
2.3	The estimation approach . . . . .	39
2.3.1	The true-score measurement error model . . . . .	39
2.3.2	Designs for estimating measurement quality . . . . .	43
2.3.3	Contamination models and other measurement models . .	52
2.3.4	Administrative data . . . . .	53
2.4	Models for data editing . . . . .	56
2.5	Conclusion . . . . .	57
<b>3</b>	<b>Algorithms for Correcting Sign Errors and Rounding Errors in Business Survey Data</b>	<b>59</b>
3.1	Introduction . . . . .	59
3.2	Current approach at Statistics Netherlands . . . . .	62
3.3	Sign errors . . . . .	63

---

## TABLE OF CONTENTS

---

3.3.1	The profit-and-loss account . . . . .	63
3.3.2	Sign errors and interchanged revenues and costs . . . . .	64
3.3.3	A binary linear programming problem . . . . .	68
3.3.4	Allowing for rounding errors . . . . .	69
3.3.5	Summary . . . . .	70
3.4	Rounding errors . . . . .	70
3.4.1	Introduction . . . . .	70
3.4.2	Matrix theory . . . . .	71
3.4.3	The scapegoat algorithm . . . . .	72
3.4.4	A real-world application . . . . .	81
3.5	Application to the Netherlands' Structural Business Statistics of 2007	82
3.6	Conclusion . . . . .	85
	Appendix 3.A A pathological example . . . . .	86
<b>4</b>	<b>Automatic Editing with Hard and Soft Edits</b>	<b>89</b>
4.1	Introduction . . . . .	89
4.2	Background . . . . .	91
4.2.1	Edits . . . . .	91
4.2.2	The error localisation problem . . . . .	92
4.2.3	The branch-and-bound algorithm of SLICE . . . . .	94
4.3	An error localisation problem with hard and soft edits . . . . .	97
4.4	A short theory of edit failures . . . . .	98
4.4.1	Numerical data . . . . .	98
4.4.2	Categorical and mixed data . . . . .	101
4.5	An algorithm for solving the error localisation problem with hard and soft edits . . . . .	102
4.6	Example . . . . .	104
4.7	Application . . . . .	106
4.8	Conclusion . . . . .	109
	Appendix 4.A Proofs . . . . .	110
4.A.1	Proof of Theorem 4.1 . . . . .	110
4.A.2	Proof of Theorem 4.2 . . . . .	112
<b>5</b>	<b>A Generalised Fellegi-Holt Paradigm for Automatic Error Localisation</b>	<b>115</b>
5.1	Introduction . . . . .	115
5.2	Background and related work . . . . .	117
5.3	Edit operations . . . . .	119
5.4	A generalised error localisation problem . . . . .	121

## TABLE OF CONTENTS

---

5.5	Implied edits for general edit operations . . . . .	124
5.6	An error localisation algorithm . . . . .	126
5.7	Simulation study . . . . .	128
5.8	Conclusion . . . . .	132
Appendix 5.A	Fourier-Motzkin elimination . . . . .	134
Appendix 5.B	A small example . . . . .	135
<b>6</b>	<b>Estimating the Validity and Bias of Administrative and Survey Variables</b>	<b>143</b>
6.1	Introduction . . . . .	143
6.2	Methodology . . . . .	145
6.2.1	Assessing validity and intercept bias using SEMs . . . . .	145
6.2.2	Estimating an SEM . . . . .	147
6.2.3	Incorporating the audit sample . . . . .	150
6.2.4	Deriving a correction formula . . . . .	151
6.3	Application: Using VAT Turnover for the Netherlands' quarterly short-term statistics . . . . .	153
6.3.1	Introduction . . . . .	153
6.3.2	Data . . . . .	154
6.3.3	Results . . . . .	157
6.3.4	Effect on publication figures . . . . .	159
6.4	Conclusions and discussion . . . . .	160
6.4.1	Discussion of results . . . . .	160
6.4.2	Assumptions and limitations . . . . .	161
6.4.3	Potential applications . . . . .	163
Appendix 6.A	Additional methodology and results . . . . .	164
6.A.1	PML estimation for SEMs . . . . .	164
6.A.2	Missing data . . . . .	166
6.A.3	Other fit measures . . . . .	167
Appendix 6.B	Parameter estimates . . . . .	168
<b>7</b>	<b>Estimating the Quality of Business Survey Data before and after Automatic Editing</b>	<b>171</b>
7.1	Introduction . . . . .	171
7.2	Application . . . . .	173
7.2.1	Automatic editing in the Netherlands' SBS . . . . .	173
7.2.2	Data . . . . .	175
7.3	Model 1: A structural equation model . . . . .	176

## TABLE OF CONTENTS

---

7.3.1	Methodology . . . . .	176
7.3.2	Results . . . . .	179
7.4	Model 2: A finite mixture model . . . . .	180
7.4.1	Methodology . . . . .	180
7.4.2	Results . . . . .	185
7.5	Discussion . . . . .	187
7.6	Conclusion . . . . .	190
	Appendix 7.A Asymptotic standard errors for Model 2 . . . . .	191
<b>8</b>	<b>Summary, Conclusions and Discussion</b>	<b>193</b>
8.1	Summary and conclusions . . . . .	193
8.2	Potential applications . . . . .	196
8.3	Discussion . . . . .	200
	<b>Bibliography</b>	<b>205</b>
	<b>Summary</b>	<b>222</b>
	<b>Acknowledgements</b>	<b>227</b>